

Author's response. *General response to both reviewers and recommender: Please accept our sincerest thanks for the opportunity to revise this manuscript. We would also like to express our gratitude to the reviewers for their precious time and willingness to read and comment on the manuscript. Below, we address all the reviewers' suggestions and lay out the revisions made. Please note that, in order to improve readability and reduce text load, we have included only those comments that request a response.*

by Zhang Chen, 28 Sep 2022 12:49

1B. The logic, rationale, and plausibility of the proposed hypotheses, as applicable.

1. One reviewer points out that the stated goals of the research are not entirely clear. As a result, there seems to be a certain disconnection between the research goals and the specific hypotheses. Both reviewers provide useful references that can be used to further strengthen the introduction and better situate the current study in the broader literature.

Author's response: *We have significantly revised the whole Introduction section (including hypotheses), please find a more elaborated response to reviewers comment below.*

2. I would like to add that it is not entirely clear to me how certain predictions are made. For instance, hypotheses 2 and 3 are said to be based on “pilot findings (Table 2), theory, and the research literature (Table 1)”, but it is unclear to me how these different sources of information are combined to arrive at a certain prediction. For instance, attention is predicted to be related to esports performance (H2a), but the meta-analysis by Sala et al., (2018) showed a very weak correlation between gaming expertise and visual attention. Gender is predicted to be not related to esports performance (H3a), but gender is actually negatively related to performance in both titles in Pilot 2, and no study reviewed in Table 1 seems to have examined the role of gender. It would be helpful if you can more explicitly explain the rationale behind each prediction.

Author's response: *We agree that the justification of our hypotheses was not entirely clear. To make it more explicit, we have created a separate document (Appendix 4) where we state what kind of evidence or theory support/do not support our hypotheses. Please keep in mind that we have significantly reformulated all hypotheses to incorporate the feedback we have received.*

1C. The soundness and feasibility of the methodology and analysis pipeline (including statistical power analysis or alternative sampling plans where applicable).

1. One reviewer has some concerns over how the practice questions are being measured, and whether the dependent variables (i.e., ranking in a game) will be comparable between the two titles examined. It will be important to carefully consider these points, as they will have implications for the interpretation of the results.

Author's response: Based on Dr. Bonny's feedback we have revised the operationalization of the practice variable. We have conducted a small pilot (Pilot 3, $N = 44$, data and their analyses are available in Appendix 5; <https://osf.io/kcaes/>) with high-ranked players to guide our operationalization of practice and deliberate practice and have checked its clarity using a different pilot ($N = 66$, <https://osf.io/2g5ys/>). In the present form, both practice and deliberate practice contains 4 different types of activities. We have also explained how we will treat our DV and that ranking from (now) three titles (CSGO, LoL, and Fortnite) will not be included in one analysis but instead, 3 separate analyses will be conducted, each on a different data (separate analyses for CSGO, LoL, and Fortnite). However, for confirmatory analyses only data from LoL and CSGO will be used (for which we have pilot data that can lead our predictions).

2. The planned analyses may not be optimal, due to the potential issue of multicollinearity, and that the dependent variable being ordinal. If you do agree that the analyses will need to be adjusted in light of these issues, the power analysis will need to be adjusted accordingly (e.g., using an ordinal regression as the planned analysis; see some more comments on power analysis below).

Author's response: Please, find our responses to comments of Dr. Bonny about issues of collinearity and ordinal nature of our DV below.

3. One reviewer points out that the decision to use $r = .1$ as the smallest effect size of interest needs to be better justified. I also wonder how we should interpret an effect size of $r = .1$ for the different predictors here. For instance, what does it concretely mean if e.g. the predictor career length has an effect size of $r = .1$? Does $r = .1$ mean an effect is very small, regardless of which predictor is involved? I think interpreting the effect sizes in the current context (e.g. something like $r = .1$ means X extra years of playing the game will lead to an increase of one rank) will help readers better grasp the magnitude of the effects, and the rationale for using a certain smallest effect size of interest.

Author's response: Please find our response to Dr. Behnke's comment below. Based on his very thought-provoking and valuable feedback, we have significantly revised the Method section, have included two tables from Pilot data and an explanation of how large the effects could be to be also practically meaningful. Therefore, we have chosen different SESOIs and our explanation is now available in Appendix 5. Our rationale is based on the analysis of unstandardized coefficients from Pilot data, i.e. how large the change in IV should be to provide meaningful change in DV.

4. I feel the statistical inferences do not match the power analysis. For H1a-c and H4 to be corroborated, the point estimate of the effect needs to exceed $r = .1$ (with $p < .05$). My understanding of equivalence testing is that this means the 95% CI does not include $r = .1$ (otherwise one cannot claim that there is a meaningful effect size). However, the power analysis seems to be based on the classic null hypothesis testing, i.e., it tests against the null hypothesis of $r = 0$. The statistical power for corroborating H1a-c and H4 (i.e., comparing the point estimate to $r = .1$) may therefore not be 80%.

Author's response: Our (now only two) hypotheses involve two different types of statistical inferences: a) classic NHST - when an effect will be present and b) equivalence testing - when an effect will be absent. Our power analysis is indeed based on the a) classic NHST - a deviation of a single linear regression coefficient from zero. As cited by Faul et al. (2009, p. 1156) in this calculation "special F tests assessing effects of a single predictor X in multiple regression models are equivalent to two-tailed t tests of $H_0: \beta_j = 0$ ". On the other hand, our approach to equivalence

testing, the TOST procedure, uses two one-sided tests. We have clarified the specific method of equivalence testing in the Design and analysis plan section. We will apply a classic TOST approach that uses 90% CI instead of 95% (90% so that the Type 1 error rate coincides with the two one-sided tests). Thus we should have more power for option b) than for option a). Since we were unable to find any free software or R package for the computation of power analysis for equivalence testing with multiple regression analysis, we have decided to use only a classic NHST approach to power analysis. We compared the sample size requirements between NHST and equivalence testing in three most typical tests (ANOVA, t-test, correlation) in the available TOSTER package and the requirements for equivalence testing were not larger than 5% compared to NHST. Thus we will oversample our sample by an additional 10%.

5. The analyses involve multiple predictors, for each of the two titles. I think multiple comparisons may be a concern, and there may be a need to correct for it by e.g. adopting an alpha level more stringent than .05.

Author's response: We discussed the possibility of applying FWER correction very carefully before our first submission. We indeed planned to conduct two regression analyses but each conducted on a different data coming from a different population. Thus, we have decided that FWER correction does not apply to our situation. Besides, on each data coming from a different game, we will conduct two regression analyses to control for confounders and not control for colliders. Instead of applying different alpha, we have decided (also from a theoretical perspective) to use a more conservative rule - effect must be present in both regressions at the same time. However, with the change of SESOI we have decided to lower the alpha level to 0.01.

1D. Whether the clarity and degree of methodological detail is sufficient to closely replicate the proposed study procedures and analysis pipeline and to prevent undisclosed flexibility in the procedures and analyses.

1. Please provide more detailed information on the questionnaires (e.g., the items of a scale, how participants respond etc.) and cognitive tasks (e.g., the trial procedure of a task, the number of trials etc.) that you plan to use. To reduce the length of the manuscript, such information may be provided in an Appendix or the Supplemental Materials.

Author's response: As suggested by Dr. Behnke, we have expanded descriptions of intelligence, teamwork, and persistence while keeping the description of cognitive variables brief with a footnote containing a link to the webpage with full documentation. If you feel we should expand the descriptions in even more detail, we will be happy to do so.

2. To reduce researcher's degree of freedom in data analysis, please share the R code that you plan to use for the current Registered Report.

Author's response: We apologize for this mistake. The R script for confirmatory analyses is now available at the OSF repository: <https://osf.io/vbkns/>.

1E. Whether the authors have considered sufficient outcome-neutral conditions (e.g. absence of floor or ceiling effects; positive controls; other quality checks) for ensuring that the obtained results are able to test the stated hypotheses or answer the stated research question(s).

1. There seem to be different interpretations of H2e. At the moment, this hypothesis is said to be supported if at least two constructs exceed $r = .1$ in one game and null is supported in the other game. However, a certain construct may predict performance in both games, yet the effect sizes can still differ. Would this be considered as support for H2e? This would require a direct comparison between the two titles, but the issue of whether the two titles are comparable (as mentioned by one reviewer) will need to be considered.

Author's response: *This issue has been resolved with the new hypotheses.*

Other minor comments:

1. Table 1. For Thompson et al. (2013), no effect size is mentioned. It would also be useful to briefly describe what a certain effect means, as you have already done for some but not all of the included studies.

Author's response: *Now all of the effects presented in Table 1 are fully described.*

2. Table 1. Li et al. (2020) has three effect sizes. It is not clear what these three effect sizes refer to.

Author's response: *We have included information in Table 1 that these three effects represent three different conditions within the cognitive flexibility test.*

3. Above Table 1, when mentioning the results of Sala et al., 2018. "cognitive control ($\bar{r} = -.16$)". Should be $r = -.16$?

Author's response: *Indeed, thank you for pointing out this error.*

I look forward to receiving your revision in due course.

Kind regards,

Dr Zhang Chen

Reviews

Reviewed by Justin Bonny, 17 Sep 2022 19:39

Manuscript Summary

The authors propose a study investigating the relative impact of deliberate practice and psychological traits on esports player performance. They point out that prior research has not differentiated between deliberate practice and other types of accumulated esports experience when predicting performance. They further differentiate esports performance between long- and short-term metrics, suggesting the contributions of experience and psychological traits to each metric may vary. They select two esports games to recruit players from, LoL and Fortnite, to test the hypotheses.

Overall Reviewer Response

The authors rightly point out that the relative contributions of different types of experience on esports player performance remains to be examined. The inclusion of pilot study results is helpful

for gauging the feasibility of the project. However, I recommend that the authors strengthen the approach of their study by utilizing prior methods for assessing cumulative experience, especially deliberate practice, and narrowing down and refining the scope of their analysis and hypotheses to more tightly connect them to the stated goals of the study. I fully support the authors intent of investigating the development of esports player performance similar to what has previously been done for traditional sports and games such as chess.

Improving measures of accumulated experience

The validity and reliability of the proposed measures of experience, especially deliberate practice, is questionable. Prior research has commonly used different metrics when it comes to deliberate practice.

I encourage the authors to review some of the prior research by Ericsson and others who used retrospective estimates of deliberate practice. In these studies, participants are asked when they started practicing an activity and to then estimate, for each year since they started, how many hours a week they engaged in deliberate practice. These responses can then be used to estimate the cumulative amount of deliberate practice a person has with a specific activity. Do note that retrospective estimates have limitations due to memory errors (see the literature for discussion). This approach can be adapted for the proposed research with each esports experience measure: play time, deliberate practice, coached practice. This would be more in line with the expertise and deliberate practice literature and put the authors on firmer ground when explaining the measure.

Here are a couple of examples where retrospective estimates have been used to assess cumulative deliberate practice:

Ericsson, K. A., Krampe, R. T., & Tesch-Römer, C. (1993). The role of deliberate practice in the acquisition of expert performance. *Psychological review*, 100(3), 363. doi: <https://psycnet.apa.org/doi/10.1037/0033-295X.100.3.363>

Meinz, E. J., & Hambrick, D. Z. (2010). Deliberate practice is necessary but not sufficient to explain individual differences in piano sight-reading skill: The role of working memory capacity. *Psychological Science*, 21(7), 914-919. doi: <https://doi.org/10.1177/0956797610373933>

The authors need to consider how much overlap there is between the different experience measures and how they are interpreted by participants. For example, compare the current version prompts for deliberate practice and daily practice time: “Think about the past 12 months. Of all the time you spend on esports, how many hours per day is deliberate practice, i.e., gaming and non-gaming activities that need focused attention and are intended for improving specific esports skills?” versus “During those years of playing GAME NAME and/or similar games, how many hours per day have you played, on average?” Are these supposed to be mutually exclusive? Does daily hours played include deliberate practice? When revising and presenting instructions to participants, the authors need to be especially concerned with multicollinearity in their analyses. If there is too much overlap in experience measures (e.g., deliberate practice is included in daily game play) then this will be a major issue when entered into the same regression equation, making it much more difficult to detect effects and tease apart the relative contributions of highly-correlated experience predictors of performance.

Author's response: Thank you for this valuable feedback and suggestions. Based on this we have conducted another (qualitative) Pilot with high-ranked esports players (N = 44) and 2 professional esports coaches, whom we asked an open-ended question: "What are the different types of practice/training (or other activities) that they have done to advance their 'in-game' performance in esports?" Their responses and the following analysis are available in Appendix 5; <https://osf.io/kcaes/>. We have checked the clarity of items using a different pilot (N = 66, <https://osf.io/2g5ys/>) and made additional changes. Our recent measure of deliberate practice contains 4 different types of activities each, including coached practice (therefore we have removed this variable as a standalone predictor). We have also edited the operationalization of naive practice, which (for exploratory analyses) contains 4 different activities and does not include deliberate practice. For confirmatory hypothesis testing, naive practice is measured only with one item. This decision was made because we found no empirical support for the remaining three naive practice activities to improve esports success, unlike gaming experience does. Currently, there should not be an overlap between practice and deliberate practice. We also discussed two alternative operationalizations of practice, one focused on recent gaming only ("Think about the past 12 months..."), and another focused on experience throughout the entire game history ("During those years of playing GAME NAME and/or similar games,..."). We have chosen the former for the confirmatory analyses and the latter for exploratory ones. We also took into consideration the accumulated time of practice by multiplying reported training habits by years of related play.

Connecting to deliberate practice and expertise literature

The authors should make greater reference to the deliberate practice and expertise literature when justifying why psychological traits contribute to skilled performance in addition to experience. There have been recent studies suggesting that factors in addition to deliberate practice are important to the development of expert-like performance (e.g., Macnamara et al., 2014, <https://doi.org/10.1177/0956797614535810>; Hambrick et al. 2020, <https://doi.org/10.3389/fpsyg.2020.01134>). These perspectives appear to align with the stated goals of the proposed research. The authors do indeed cite some of the effects of deliberate practice from these lines of research. However, the authors focus on the effect of deliberate practice only; they should place greater emphasis on how this literature has provided evidence that other factors, such as psychological traits, in addition to deliberate practice are important for the development of skilled performance. By doing so the authors can frame the study around two competing theories in accounting for skilled performance: deliberate practice theory and the more recent theory that deliberate practice is not sufficient alone in developing skilled performance.

The specific studies that these articles reference may also be useful for identifying prior methods and analyses that could be used as frame of reference for the present research.

Author's response: We are very thankful for your suggestions. Based on that we have revised the Introduction and included also a brief summary of other factors mentioned in referenced studies. Regarding the last note, namely to what degree DP is sufficient for developing skilled performance, we did not include in our confirmatory design (which is already quite complex) but we will certainly discuss the effect ratio between practice and other variables with related theories after the

results are in, at Stage 2. Based on these results, a new design could then be developed to be tested with competing hypotheses in the future.

Connecting hypotheses back to study goals and stated gaps in research

It seems that too much emphasis has been placed on the pilot study results at the expense of the stated research goals. These goals, as listed in the current version of the manuscript, are as follows:

“In the present confirmatory study, our goal is test whether deliberate practice theory, which has successfully been applied to other sports earlier, can also predict high esports performance.”

(abstract)

“Along these events, a relevant research question has emerged: what skills and attributes are needed to become a successful esports player? This is our research question.” (first paragraph of introduction)

“In the present study, our goal is to test if the deliberate practice theory of performance development applies to esports, and how other psychological and environmental components might be relevant for esports performance, too.” (second paragraph of introduction)

To me, in synthesizing these statements, the goal of the study is to investigate the relative impact of deliberate practice and psychological traits on the development of esports player performance. However, some of the stated hypotheses do not seem relevant to the stated study goals and are instead the result of the pilot study (e.g., effect of gender, age, teamwork and physical training). If my synthesized goal is indeed what the authors intend to examine with their study, I encourage them to revise and condense their hypotheses to be more specifically aligned with the study goal. They should certainly be informed by the pilot study, but more connected to the goal. For example, the authors may hypothesize that cognitive trait measures (attention, reaction time) will be a stronger predictor of skilled performance than deliberate practice.

However, if I am incorrect in my interpretation of the study goal, the authors should revise the introduction to focus on the pilot study-generated hypotheses and identify the theories that support them more clearly.

Author's response: *Thank you for pointing out this discrepancy. We acknowledge that we have omitted to state demographic factors in the research goal in order to align it with the hypotheses. We have edited our research goal accordingly.*

We have also reformulated our research question with this comment in mind. In the following second paragraph of introduction, we now state more specifically "In the present study, our goal is to test if the deliberate practice theory of performance development applies to esports, and how other psychological and environmental components might be relevant for esports performance, too." the role of what kind of factors we are going to verify.

Also to make justification of our hypotheses more explicit and clear, we have created a separate document (Appendix 4) where we state what kind of knowledge/research is supporting our hypotheses.

Due to the absence of robust previous evidence or a formalized theory, we have refrained from making explicit hypotheses about the order of importance of specific factors or about the size of the differences between their effect sizes.

Clarifying Skilled Performance

In the introduction the authors compare short- versus long-term skilled performance. In doing so I was expecting this to be a component of the study design: including measures of short-term and long-term performance. However, only a measure of long-term skilled performance is included in the methods section. The authors should either include a measure of short- and long-term performance or remove this distinction in the introduction. Furthermore, the authors need to justify how their measure of skilled performance (highest rank in [game]) is a valid and reliable indicator of long-term skilled performance.

Author's response: *We have included the paragraph explaining the difference between short-term and long-term performance purposefully for two reasons: 1) we are not aware of such an explicit distinction in the existing literature and we consider it important to explain that our study deals only with one specific kind of performance, and 2) to justify the inclusion of published literature dealing with only long-term performance as being distinct from short-term one. However, we agree that our position and focus on long-term performance need to be clarified. We have edited the Introduction section accordingly. The inclusion of short-term performance would definitely be interesting and enriching, however not easily compatible with the existing setup of our study. We have also included the description of our dependent variables in the Method/measures section, where we explain how ranks (our DV) represent skilled performance.*

Justifying Inclusion of Two Esports Titles

It is unclear why two esports games are included in the study. Are the authors using this as a way to generalize across players of different esports games or to compare and contrast the relative contributions across different games? Please include additional justification in the introduction. Additionally, it would be helpful to include a comparison of the titles based on included game mechanics rather than describing them as being from different genres.

A concern about including two different esports titles is the equivalency of skilled performance metrics. The stated measure of long-term skilled performance is “In the past 12 months, what is your highest rank in [game]?” with the indicated rank presumably being converted into a number (this needs to be clarified) with higher numbers indicating higher skilled performance. However, in doing so, the authors are assuming that, for example, a rank of 5 in LoL is equivalent to a rank of 5 in Fortnite. But is that actually the case? The authors need to justify the use of this operational definition of skilled-performance across both esports titles. If these ranks are not equivalent across titles this could lead to serious issues in statistical analyses. For example, the authors could observe a between-title difference of deliberate practice when predicting skilled performance: would this be due to an actual difference in impact of deliberate practice (i.e. it is more important for LoL than Fortnite) or due to dependent variable of rank for one title not being equivalent to the other?

Author's response: *We apologize for not being clear in explaining the role of 2 different game titles. Since we have significantly revised our methodology, with lower demands on sample sizes we decided to also include a third sample, players of Counter-Strike: Global Offensive. For confirmatory analyses, we will use samples of LoL and CSGO players (we have pilot data that can*

lead our predictions) and for exploratory analyses, we will use the sample of Fortnite players (for which we do not have pilot data).

Indeed, we have included two esports titles to both extend the generalizability of our results and to compare the relative contributions across different games. However, we are not using the same operationalization of game performance across the three (or previously two) esports games. Thus the problem of non-equivalence of two different metrics does not apply to our analysis. Instead of treating players of the two games as coming from the same population (using 1 sample) we test our hypotheses on 2 different samples (assuming the existence of three different populations – different games require different abilities, have a different distribution of ranking, etc.) with different operationalization in each game (ELO based rating with 18 divisions in CSGO, MMR based rating with 27 ranks in LoL, and Hype points based rating with 10 ranks in Fortnite). Now, we have made the use of three samples and different operationalizations clearer throughout the manuscript. We have also supported justification for the inclusion of three games in the Introduction (fourth paragraph) and we have also included a brief description of game mechanics related to each of the three esports games (Methods section/Design).

Refining Planned Analyses

Is the dependent variable, skilled performance, an ordinal measure? If so, the authors should consider whether a generalized linear model is required for analysis instead of a linear regression and provide justification.

Author's response: Based on this valuable feedback, we have updated the way we plan to analyze data.

We agree that our dependent variable, skilled performance could be considered an ordinal variable. Level of skill between each of the three (out of the 18 divisions in CSGO, 27 ranks in LoL, and 10 ranks in Fortnite) adjacent ranks (e.g. the difference between Grandmaster vs Master and Diamond vs Platinum) does not have to be the same. On the other hand, all three rank systems are based on a number of points (ELO rating in the case of CSGO, Matchmaking Rating in the case of LoL, and Hype points in the case of Fortnite) – thus truly interval variables. In psychology, a vast majority of truly ordinal variables are treated as an interval for different reasons - we agree that many times incorrectly. Some research (Li, 2016; Rhemtulla et al., 2012) have however provided evidence that ordinal variables could be in some instances (e.g. number of categories, distribution of data etc.) modeled as continuous.

We also had to make a pragmatic decision in the choice of statistical procedure – to find a solution in R that can test all of our hypotheses (to test both presence and absence of effects higher than our SESOI). For that pragmatic reason, we decided to use Robust linear regression (High Breakdown-Point and High Efficiency robust estimates; Yohai, 1987) instead of e.g. Ordinal logistic regression. Robust linear regression with MM estimation should provide more accurate estimates than OLS we originally proposed.

The entry of all variables into a single regression model makes it very likely that multicollinearity will be a problem, especially for measures of experience (but see earlier comment about possible ways to reduce this). The authors need to specify how will this be addressed.

Author's response: We have discussed the problem of collinearity during our Pilot studies, when including predictors that could be potentially highly correlated, e.g., different practice times or cognitive measures. One of the solutions that literature is commonly recommending is dropping one of the correlating predictors. This would be justifiable when the original model is considered misspecified. However, we consider each of the predictors included in the model as meaningful and tapping on a different construct, e.g. coached practice to be meaningfully different than standard daily practice or speed of decision-making measure different construct than reaction time. For this reason, in accordance with the opinion of Vanhove, 2021 (doi:10.15626/MP.2021.2548) that collinearity does not represent a problem for regression analysis, we have decided not to drop any of the correlated predictors. Apart from our decision, we did not observe collinearity in the pilot data. Moreover, we will supplement our results with zero-order estimates (i.e., DV is predicted only by one IV at time).

Are the authors predicting that psychological traits will moderate / mediate the impact of deliberate practice and experience on performance? This would be in line with previous studies that have observed such interactions. This needs to be stated more clearly in the analysis plan.

Author's response: We originally considered testing one moderation and one mediation effect, but after considering the mixed evidence in the previous literature, we decided to refrain from these analyses.

The authors suggest interaction effects in H2 but the sample size is based on main effects observed in the pilot study. The sample sizes need to be re-estimated to detect the smallest predicted interaction effects. It appears that the pilot study can be used to estimate the effect size of such interactions for use in a revised power analysis

Author's response: We are sorry for not being clear in formulating H2. We did not want to hypothesize an interaction effect (1 regression analysis with moderating effect of game genre on the relationship between esports performance and its predictors) but a different composition of predictors in multiple game genres (in our case 3 different regressions, each conducted on data from different populations). Thus, instead of interaction (1 regression), we plan to conduct 2 (2 in each game title) regressions to test the generalizability (applicability of our findings in different contexts/populations) of our results. We have explained our intention and H2 in more detail in the Design section. Please keep in mind that we have significantly reformulated all hypotheses to incorporate the feedback we have received.

Reviewed by Maciej Behnke, 02 Sep 2022 08:02

This manuscript proposes the examination of the psychological predictors of long-term esports success. The topic is interesting, and I enjoyed reading the manuscript. Please find my comments below.

Suggestions to consider:

1. **Title.** The study will examine long-term esports success. Do you want to specify the type of esports success in the title?

Author's response: Great point. Thank you. We have edited the title name.

2. **Power.** I'm aware that the default for power is 0.80, but some journals require a power of .90 or .95 (NHB). Do you want to keep the aimed power of .80 and limit the scope of journals willing to publish your study? I'm aware that increasing the power to .95 would increase your sample to 705 participants in each group, but maybe it is something worth considering.

Author's response: *We will gather our data from Prolific participants. To make sure that we can achieve our sample size on Prolific, we have already prescreened 10 000 of their participants (players of digital games) to select only those who meet our inclusion criteria (play ranked games in a specific game). Based on our experiences from collecting data for Pilot studies we have found that response rates (proportion of participants accepting our call in less than two weeks) were approx. 50% (despite the fact that we have chosen a decent compensation). A limiting factor here may be the time required to complete our survey which is around 40 minutes. In our original plan and SESOI of $r = .1$, we considered it unrealistic or too risky to promise to have such a large sample. Since we have significantly revised the whole methodology of our research, with the change of SESOI to (also practically meaningful) $r = .2$, we will be able to collect data to achieve 0.95 power for the specified ES.*

Comments:

L155 – In the introduction, you presented the meta-analysis and systematic reviews related to cognitive skill, whereas the theoretical models mentioned earlier also identified other dimensions of sports competencies (e.g., psychic or mental abilities). I would suggest adding an existing systematic review on the effects of stress on gaming performance in the introduction (Leis, & Lautenbach, 2020).

Leis, O., & Lautenbach, F. (2020). Psychological and physiological stress in non-competitive and competitive esports settings: A systematic review. *Psychology of sport and exercise*, 51, 101738.

Author's response: *Thank you for your suggestion. We have included the results of this systematic review in the Introduction section.*

L168 – Here, I believe the quantitative study by Leis et al. (2022) might be helpful.

Leis, O., Lautenbach, F., Birch, P. D., & Elbe, A. M. (2022). Stressors, associated responses, and coping strategies in professional esports players: A qualitative study. *International Journal of Esports*, 3(3).

Author's response: *Thank you very much for this recommendation. We have decided, however, not to cite this study, since stress control was already present and mentioned as a component of Nagorsky and Wiemeyer (2020) theory and the study of Leis et al. (2022) does not directly relate coping of stress to esports performance.*

Table 1: Please explain the abbreviation used in the Table (e.g., MMR).

Author's response: *We are sorry for the omission of this information. We now explain all of the abbreviations used in Table 1.*

Table 1: Although I'm delighted to see my work in table 1 (Behnke et al., 2020), I'm not sure whether it fits here. The table presents the relationships between long-term esports success and environmental, psychological, and other factors. In our studies (also in Behnke et al., 2022), we examined factors (e.g., emotions and stress) that influence short-term esports success.

Author's response: Thank you for pointing this out. In the former version of this Table, we have included both short-term and long-term performance. By mistake, the row referring to your study was left in the table. Now we have removed this reference.

L254: Can you explain why $r = 10$ was chosen as SESOI? It might be interesting if you could back up your decision with some data. Ps. Please explain what SESOI means. Although it might be known to the preregistration community, most researchers will not be familiar with this term.

Author's response: Our new SESOIs are based on a critical discussion of standardized and unstandardized regression coefficients found in Pilot 2 - how large the change in IV should be to provide meaningful change in DV. We tried to choose the smallest effect that could still have a motivational potential - is it still possible and/or worth spending so much effort to increase the performance by just one rank? We acknowledge that the choice of SESOI in our research can be debated in future studies. We have tried to make our justification of the SESOI as clear and transparent as possible. Therefore, in Appendix 5 we describe the rationale of choosing selected SESOI for each predictor.

By screening psychological literature, we can see that our SESOI resemble effect sizes typical for psychological literature ("the 25th, 50th, and 75th percentiles corresponded to correlation coefficient values of 0.12, 0.24, and 0.41" - Lovakov & Agadullina, 2021, [10.1002/ejsp.2752](https://doi.org/10.1002/ejsp.2752); median $r = 0.16$ from preregistered studies, Schäfer & Schwarz, 2019, [10.3389/fpsyg.2019.00813](https://doi.org/10.3389/fpsyg.2019.00813)), and an esports related literature with results summarized in Table 1.

Again, we want to emphasize that whether an effect size of 0.2 or 0.3 has practical meaning and impact on esports performance will ultimately always depend on the individual.

L300 – I believe that you can provide more details about the measures, especially about the psychological questionnaires. As in principle, you should not make any changes in the methods section after IPA; I believe you should describe measures in greater detail.

Author's response: We concur that the description of the measures was probably too brief. We have expanded descriptions of intelligence, teamwork, and persistence while keeping the description of cognitive variables brief with a footnote containing a link to the webpage with full documentation. If you feel we should expand the descriptions in even more detail, we will be happy to do so.

L315: Can you provide the code for your analysis in the supplementary materials?

Author's response: We apologize for this mistake. The R script for confirmatory analyses is now available at the OSF repository: <https://osf.io/vbkns/>.

Minor point: L46, I believe the – “goal is TO test”.

Author's response: Thank you for pointing to this mistake. Corrected.